Developing a Transformer-Based Sensor Verification Service for a Flood Intelligence Application, floodwatch.io

Andrew Ma, Abhir Karande, N. Rich Nguyen University of Virginia

Introduction

The FloodWatch system relies on accurate sensor readings for producing granular flood intelligence. Anomalistic readings could have drastic consequences

Anomalistic Readings often caused by[1]
- Physical Disturbances
- Poor Signal Strength
- Low Battery Life
- Old Sensors

Prior sensor anomaly detection work is focused on analyzing a sensor individually
Challenges
- Extreme weather (i.e. heavy rainfall, extreme heat)
- Missed readings (i.e. sensor doesn't detect rain)

Corroboration Mechanisms

We propose adding corroboration mechanisms to improve anomaly detection (Fig. 2)

1. Pair sensors based on location (Fig 3.)
   - Utilize minimum weight matching to group closest clusters
   - Analyze the difference in readings between pairs (Fig. 1)
2. In the case of an anomaly in the pairs, then compare sensor readings to WeatherAPI data

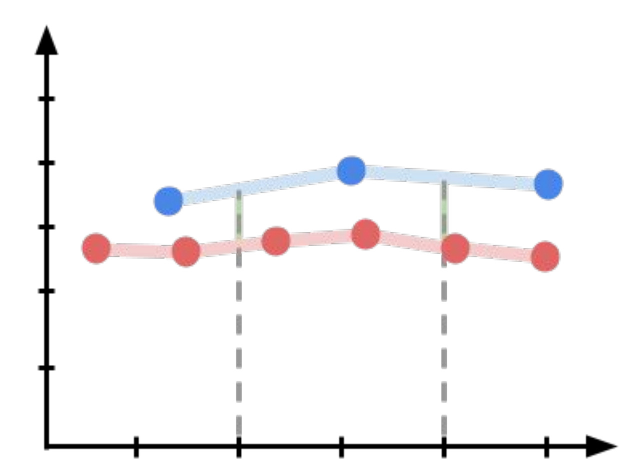

Figure 1. Sensor readings are combined by linearly interpolating the readings, and taking the differences at specified intervals (highlighted green)

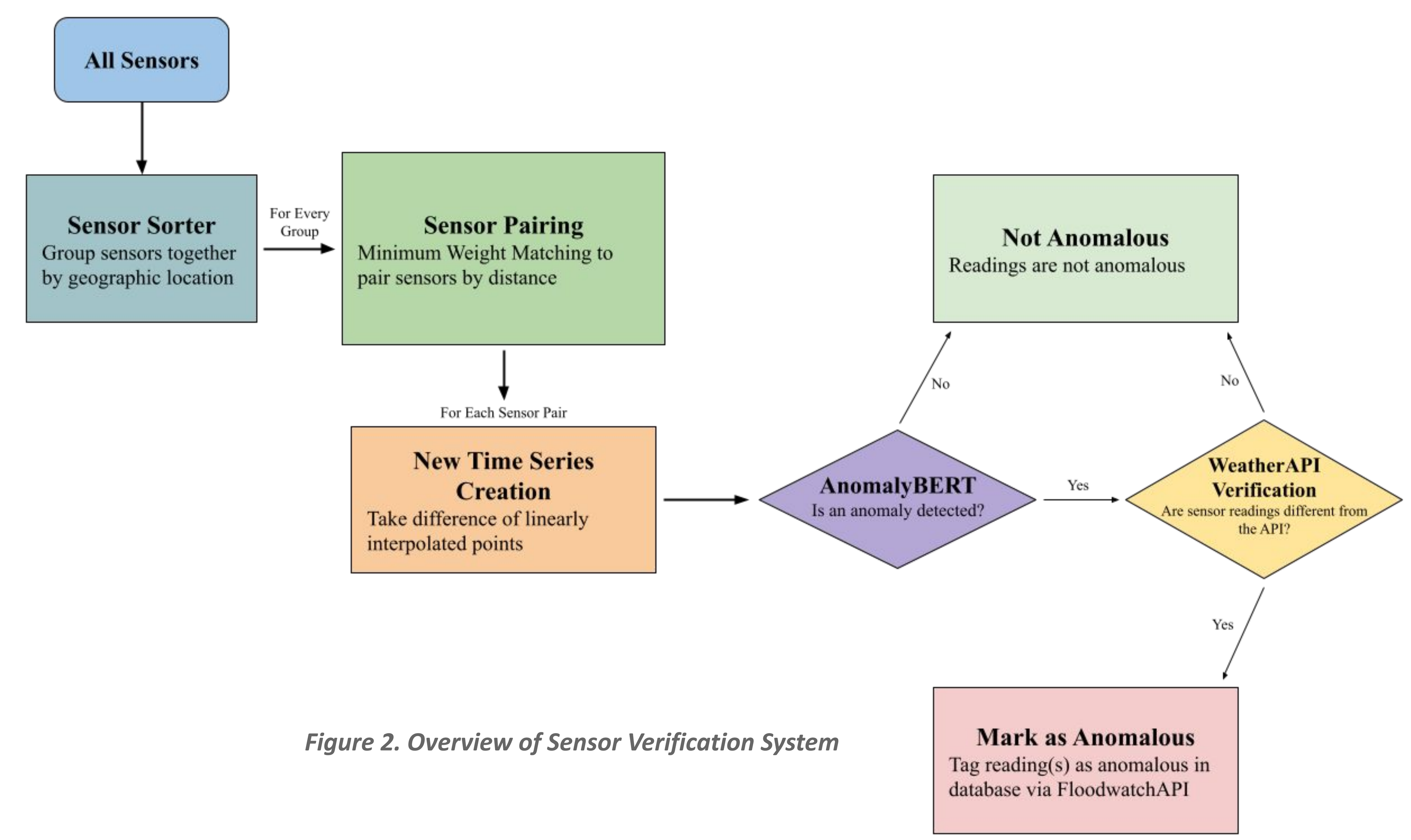

Figure 2. Overview of Sensor Verification System

Figure 3. Sensor pairs for deployed sensors in Da Nang, Vietnam. Pairs are created so that total distance between pairs is minimized.[2, 3]

Data

Mock data was used to test the effectiveness of this system. 2,000 hours of Ho Chi Minh City weather data was gathered from WeatherAPI[3]. Different types of synthetic noise was then added to simulate 8 different sensor situations.
Different types of sensors:
- Normal data with noise
- Normal data with constant offset (Fig. 4)
- Oscillating offset

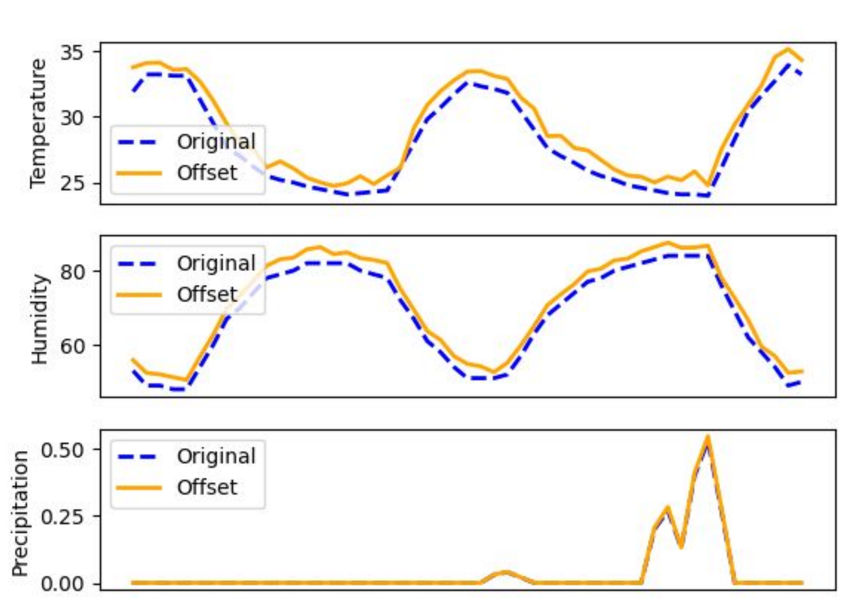

Figure 4. Example of offset data used in mock data

Afterwards, anomalies were manually added and labeled to the sensor data. Both point and contextual anomalies were included (Fig. 5).

AnomalyBERT

Transformer architecture that makes use of self-supervised learning
Utilizes data degradation to generate synthetic outliers
- Selects random interval and alters data on that interval through one of four degradation methods
- Soft replacement, Uniform replacement, Peak noise, and Length adjustment[5]

Metrics
Exact Values:
- precision: 0.833 | recall: 0.741 | F1-score: 0.784

Point Adjusted Values:
- precision: 0.952 | recall: 0.741 | F1-score: 0.833

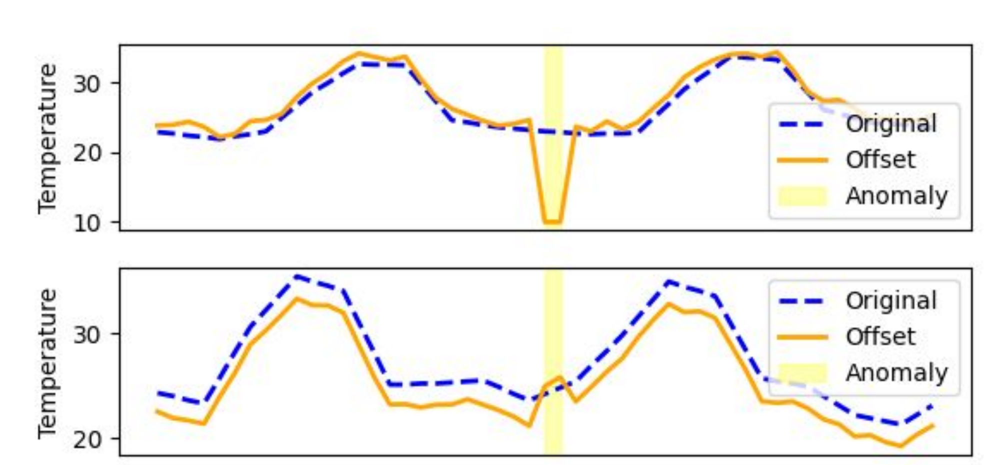

Figure 5. Examples of point anomaly (top) and contextual anomaly (bottom)

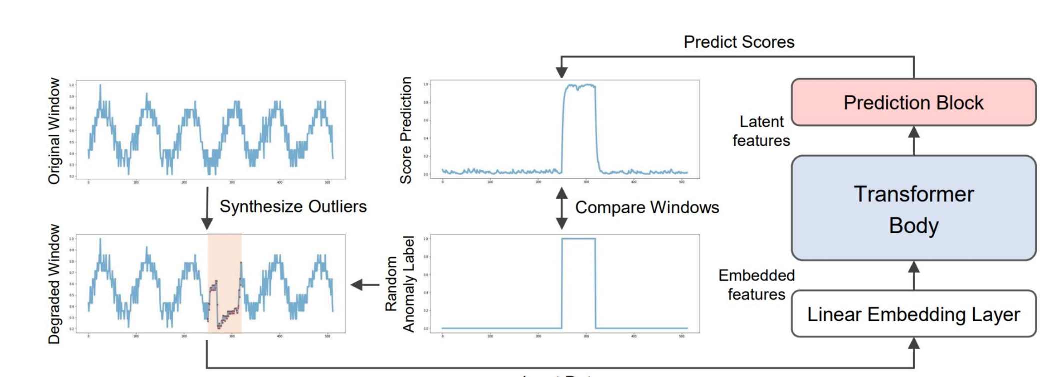

Figure 6. AnomalyBERT's self-supervised training framework that utilizes data degradation[5]

Automated Service

The sensor verification service runs daily, handling
- Sensor pair matching
- Calling endpoint for batch inferencing
- WeatherAPI verification

Serverless Inference

To be used with our automated service, a trained AnomalyBERT model is deployed to an inference endpoint
- Serverless architectures removes need for provisioning and maintenance of server
- Low-Latency batch inferencing
- AWS Sagemaker

Anticipated Results

- Ability to detect specific types of anomalies and attribute it to a specific causal factor
- Automated verification of full sensor network

Future Work

- Dynamic sensor matching
- Continual/online-adaptive learning for AnomalyBERT

Sources

Nguyen, L. A., Kiet, P. T., Lee, S., Yeo, H., & Son, Y. (2023). Comprehensive Survey of Sensor Data Verification in Internet of Things. IEEE Access.
OpenStreetMap, www.openstreetmap.org
Build Dynamic Maps with Mapbox, www.mapbox.com/about/maps/
WeatherAPI, www.weatherapi.com
Jeong, Y., Yang, E., Ryu, J. H., Park, I., & Kang, M. (2023). AnomalyBERT: Self-Supervised Transformer for Time Series Anomaly Detection using Data Degradation Scheme. arXiv preprint arXiv:2305.04468.

UNIVERSITY of VIRGINIA